Visual Tracking Algorithm using Pixel-Pair Feature

Kenji NISHIDA
Neuroscience Research Institute
National Institute of Advanced
Industrial Science and Technology(AIST)
Central 2, 1-1-1 Umezono,
Tsukuba, IBARAKI 305-8568 JAPAN
kenji.nishida@aist.go.jp

Takio Kurita*
Neuroscience Research Institute
National Institute of Advanced
Industrial Science and Technology(AIST)
Central 2, 1-1-1 Umezono,
Tsukuba, IBARAKI 305-8568 JAPAN
takio-kurita@aist.go.jp

Yasuo Ogiuchi
Information & Communication Laboratories
Sumitomo Electric Industries Ltd.
1-1-3, Shimaya, Konohana-ku
Osaka, 554-0024 Japan
ogiuchi-yasuo@sei.co.jp

Masakatsu Higashikubo
Information & Communication Laboratories
Sumitomo Electric Industries Ltd.
1-1-3, Shimaya, Konohana-ku
Osaka, 554-0024 Japan
higashikubo@sei.co.jp

Abstract

A novel visual tracking algorithm is proposed in this paper. The algorithm uses pixel-pair features to discriminate between an image patch with an object in the correct position and image patches with an object in an incorrect position. The pixel-pair feature is considered to be robust for the illumination change, and also is robust for partial occlusion when appropriate features are selected in every video frame. The tracking precision for a deforming object (skier) is examined and also the occlusion detection method is described.

1 Introduction

Visual object tracking is one of the most important issue in computer vision, because of its usefulness for many applications such as surveillance systems and intelligent transport systems. The object tracking algorithms can be classified to four well-known approaches. The first is based on the background subtraction algorithm [13]. In this approach, moving objects are extracted from the background. However, the performance of background estimation is degraded when the movement of objects are small, and it requires an appropriate illumination condition.

The second approach can be called the feature-based tracking algorithm [6, 5, 7]. Salient features such as corner features are individually extracted and tracked and are grouped in this approach. This approach is robust to changes in the illumination condition. However, the precision of the object location and dimension is affected by the difficulties in feature grouping.

The third approach is called the mean-shift[8, 9] approach, in which local features (such as color histograms) of pixels corresponding to the object are followed. The mean-shift approach enables robust and high-speed object tracking, if a local feature that can be used to discriminate between the object and the background exists. However, it is difficult to discriminate between nearby objects with similar colors and to adopt this method for gray-scale images.

The fourth (and final) approach can be classified as a discriminative tracking approach. Avidan[10] redefined the tracking problem as the problem of classifying (or discriminating between) objects and the background. In this approach, features are extracted from both the objects and the background; then, a classifier is trained to classify (discriminate between) the object and the background. Hidaka[11] employed rectangle feature for their classifier and tracked objects by maximizing the classification score. Grabner[12] trained a classifier to discriminate an image patch with an object in the correct position and image patches with objects in the in-correct position, thereby, the position of the

*current address: Hiroshima University (http://home.hiroshima-u.ac.jp/tkurita/)

object could be estimated in higher precision. While this approach enables stable and robust object tracking, a large number of computations are necessary. The approach of Collins[4] and Mahadevan[14] is classified as an approach of this type, but they selected discriminative features instead of training classifiers.

Object trackers must cope with the appearance change (illumination change, deformation etc.), thus most of the trackers update feature set while tracking a particular object. Although updating feature set improves the tracking performance, it may affects the precision (or stability) of tracking such as drifting[15]. Therefore, Grabner[15] introduced on-line boosting to update feature weights to attain a compatibility of adaptation and stability of tracking classifiers. Woodley[16] employed discriminative feature selection using a local generative model to cope with appearance change while maintaining the proximity to a static appearance model.

In this paper, we propose a tracking algorithm for gray-scale images that reserves *good* pixel-pair features by discarding insufficient pixel-pair features and refilling new pixel-pair features from randomly selected pixel-pairs in every video frame. The pixel-pair feature is determined by a relative difference in the intensities of two pixels[1, 2, 3]; therefore, it is considered to help realize robustness to illumination changes. Our algorithm is applied to the traffic surveillance system[17] and showed good tracking performance for fixed shape objects such as cars, and also showed the good performance for partial occlusion without taking steps for occlusion detection. However, to obtain a general surveillance systems, the applicability to deforming objects such as human, and occlusion detection are important.

2 Tracking Algorithm using Pixel-Pair Feature

2.1 Tracking Procedure

We define a tracking problem as a classification problem of obtaining an image patch that contains the object in the correct position from a new image frame. A tracking procedure is briefly illustrated in figure 1. For the t_{th} frame, the image frame $\boldsymbol{V}_t$ and a vehicle position (and scale) $\boldsymbol{L}_t$ are obtained from the $(t-1)_{th}$ frame (the initial position is given by the vehicle detector). Our tracking system can be used to crop a positive (correct) image patch $\boldsymbol{I}_t$ using $\boldsymbol{V}_t$ and $\boldsymbol{L}_t$; then, F false (incorrect) image patches $\boldsymbol{J}_t^1, \ldots, \boldsymbol{J}_t^F$ surrounding $\boldsymbol{L}_t$ are cropped. Next, the features for discriminating between $\boldsymbol{I}_t$ and $\boldsymbol{J}_t$s are extracted, instead of training a classifier. Finally, a search for an image patch $\boldsymbol{I}_{t+1}$, which is the most similar to the positive (correct) image patch $\boldsymbol{I}_t$, is carried out from the next frame $\boldsymbol{V}_{t+1}$.

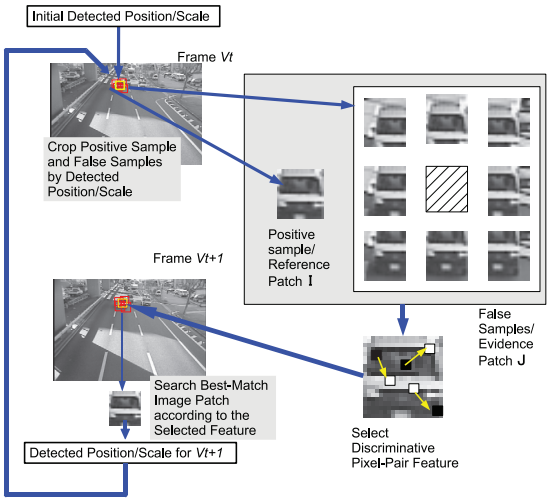

Figure 1. Tracking procedure

2.2 Pixel-Pair Feature

We adopt pixel-pair feature for our tracking algorithm, since it is powerful for estimating the similarity between one reference image (template) and many evidential images (this property is common in similar features described in [1, 2, 3]). The pixel-pair feature is an extension of the statistical reach feature (SRF)[1] in which the restriction on the distance between pixel pairs is removed. The definition of the pixel-pair feature and the similarity index $c(I, J)$ of a given pair of reference image I and evidential image J of the same size are described as follows (figure 2). Suppose the size of the input images is $W \times H$. Let grid Γ represent a set of pixel coordinates in the images I and J, such as $\Gamma := \{(i, j) | i = 1, \ldots, W, j = 1, \ldots, H\}$. We regard the image of size $W \times H$ as an intensity function defined on Γ. For an arbitrary pair (p, q) of grid points in Γ, we define the value $ppf(p \succ q; T_p)$ as follows:

$$ppf(p \succ q; T_p) := \begin{cases} 1 & \boldsymbol{I}(p) - \boldsymbol{I}(q) \geq T_p \\ -1 & \boldsymbol{I}(p) - \boldsymbol{I}(q) \leq -T_p \\ \phi & otherwise \end{cases} \quad (1)$$

Here, $T_p (> 0)$ is the threshold of the intensity difference. We adopt the grid-point pair (p, q) as a feature when $ppf(p \succ q; T_p) \neq \Phi$. Hereafter, we write $ppf(p \succ q)$ rather than $ppf(p \succ q; T_p)$, unless there is any ambiguity.

We limit the number of pixel-pair feature to N by selecting a set of pairs $(\boldsymbol{p}, \boldsymbol{q})$ with selection policy s,

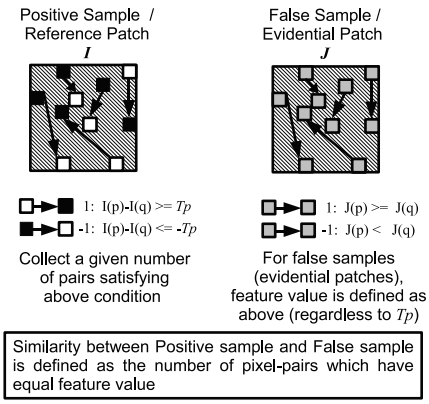

Figure 2. Pixel-pair Features

we denote a pixel-pair feature set RP_s as follows;

$$RP_s(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{I}, T_p, N) := \{ppf(\boldsymbol{p} \succ \boldsymbol{q}) \neq \Phi\}, \quad (2)$$

where $\{\boldsymbol{p}, \boldsymbol{q} \in \Gamma \times \Gamma\}$, $\boldsymbol{p} = \{p_1, \ldots, p_N\}$, and $\boldsymbol{q} = \{q_1, \ldots, q_N\}$.

We define the incremental sign $b(\boldsymbol{p} > \boldsymbol{q})$ for the evidential image $\boldsymbol{J}$ for computing the similarity between images $\boldsymbol{I}$ and $\boldsymbol{J}$ as follows:

$$b(\boldsymbol{p} \succ \boldsymbol{q}) := \begin{cases} 1 & \boldsymbol{J}(p) \geq \boldsymbol{J}(q) \\ -1 & otherwise \end{cases} \quad (3)$$

For a pixel pair $(p, q) \in RP_s$, a single-pair similarity $r(p, q, \boldsymbol{J})$ is defined as follows:

$$r(p, q, \boldsymbol{J}) = \{ppf(p \succ q) = b(p \succ q)\}. \quad (4)$$

$c_s(\boldsymbol{I}, \boldsymbol{J}, RP_s)$, the similarity index between image I and J, measured by using a pixel-pair feature set RP_s is defined as follows:

$$c_s(\boldsymbol{I}, \boldsymbol{J}, RP_s) = \frac{\sum_{(p,q) \in RP_s} r(p, q, \boldsymbol{J})}{|RP_s|} \quad (5)$$

At the first video frame for tracking an object, pixel-pairs are selected according to the single-pair similarities (lower similarity pairs are selected) from randomly selected pixel-pair feature set. At subsequent video frames, previously selected pairs are examined to have valid ppf value described in expression (1). The pixel-pairs that do not have valid ppf value are discarded, and refilled from randomly selected pixel-pairs. If a pixel-pair is reserved through many video frames, the pixel-pair is considered a *good* feature for tracking the object, and the collection of long-life pixel-pairs may represent the key description of the object.

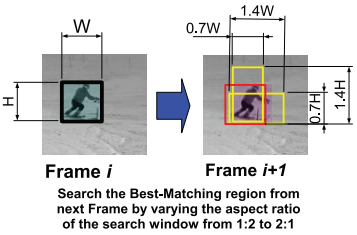

Figure 3. Tracking Deforming Objects

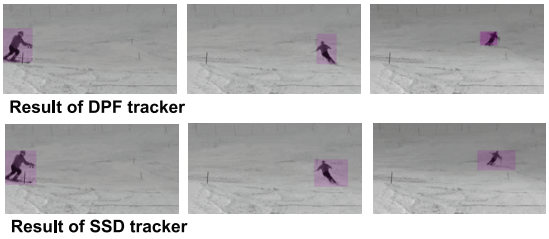

Figure 4. Tracking Result for a Skier

3 Experiment

3.1 Tracking a Deforming Object

Tracking a skier is considered more difficult than tracking a pedestrian, since the body (torso) leans in every turn. We employ deforming model that changes the vertical/horizontal ratio of rectangular area for a skier (fig. 3), instead of introducing complicated deforming model such as articulated body etc.

To validate our algorithm based on the discriminative pixel-pair feature (DPF tracker), the result is compared with the result of tracking algorithm based on the least sum of the squared difference (SSD tracker). Figure 4 shows that the DPF tracker attains a precise tracking, while the shape and scale error in SSD tracker are af-

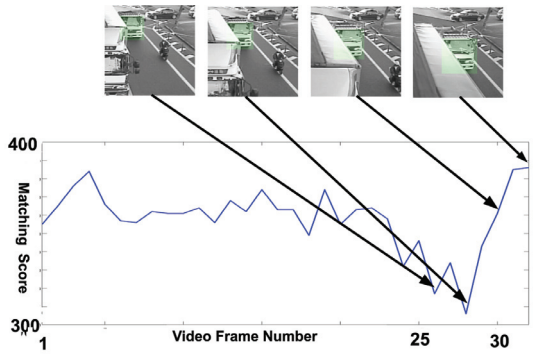

Figure 5. Matching Score for Occlusion

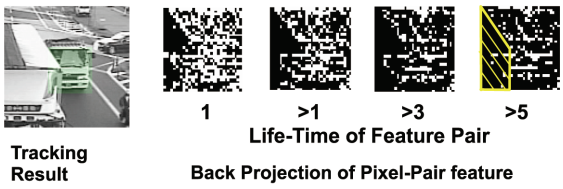

Figure 6. Back Proj. for Pixel-Pair Feature

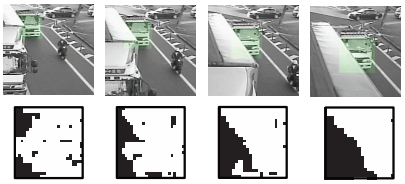

Figure 7. Result of Occlusion Detection

fected by deformation of the skier.

3.2 Occlusion Detection

Figure 5 shoes the matching score for occluded vehicle. Occlusion begins from 24th video frame and continues to the 32nd video frame. The matching score degrades in the beginning of occlusion, but the score is recovered even though the occlusion is continued. Since DPF tracker updates features to obtain a highest matching score for a tracking object in every video frame, the score does not show significant deterioration in case of partial occlusion. However, if the region where pixel-pairs are extracted is occluded between two consecutive video frames, the pixel-pairs would lose there matching score, thus they would be discarded in the next video frame. Therefore, back projection of pixel-pair with long *life-time* (reserved for several video frames) should indicate the occluded region (fig.6). Figure 7 shows the detected occluded region from the back projection of pixel-pairs reserved more than 5 consecutive video frames.

4 Conclusion

We proposed a novel algorithm for object tracking. In the proposed method, pixel-pair features that are considered to be suitable to find a most similar image patch from many candidates. The proposed method showed a high robustness for the object deformation with a simple variation of vertical/horizontal ratio of image patches. The detection of occluded region according to the *life-time* of pixel-pairs is also examined, and the occluded region can be approximated according to the back projection of pixel-pairs with long *life-time*. We are now developing a multi-object tracking system based on the proposed algorithm.

References

[1] R.Ozaki, et al., "Statistical Reach Feature Method and Its Application to Template Matching", in *Proc MVA 2009*, pp.174-177, 2009.

[2] S. Kaneko, et al., "Robust image registration by increment sign correlation",*Pattern Recognition*, vol.35, no.10, pp.2223-2234, 2002.

[3] M.Özuysal, et al., "Fast Keypoint Recognition in Ten Lines of Code", in *Proc. CVPR 2007*, 2007.

[4] R.T.Collins, et al., "Online Selection of Discriminative Tracking Features", in *IEEE PAMI*, Vol.27, No.10, pp.1631-1643, 2005.

[5] B.Coifman, et al., "A real-time computer vision system for vehiclr tracking and traffic surveillance", *Transportation Research Part C*, No.6, pp.271-288, 1998.

[6] D.Beymer, et al., "A Real-Time Computer Vision System for Measuring Traffic Parameters", in *proc. IEEE CVPR*, pp.495-501, 1997.

[7] Z.Kim, J.Malik, "Fast Vehicle Detection with Probabilistic Feature Grouping and its Application ot Vehicle Tracking", in *Proc. ICCV*, pp.524-531 2003.

[8] D.Comeniciu, P.Meer, "MeamShift: A Robust Approach Toward Feature Space Analysis", *IEEE PAMI*, Vol.24, No.5, pp.603-619, May, 2002.

[9] D.Comaniciu, et al., "Real-time tracking of non-rigid objects using mean shift", in *proc. CVPR 2000*, Vol.2, pp.142-149, 2000.

[10] S.Avidan, "Ensemble Tracking", *IEEE PAMI*, Vol.29, No.2, pp.261,271, 2007.

[11] A.Hidaka, et al., "Object Tracking by Maximizing Classification Score of Detector Based on Rectangle Features," *IEICE Trans. on Information and Systems*, Vol.E91-D, No.8, pp.2163-2170, 2008.8.

[12] H.Grabner, M.Grabner, H.Bischof, "Real-Time Tracking via On-line Boosting", in *Proc. BMVC*, pp.47-56, 2006.

[13] D.Koller, J.Weber, J.Malik, "Robust multiple Car Tracking with Occlusion Reasoning", in *proc. ECCV*, Vol.A, pp.189-196, 1994.

[14] V.Mahadevan, N.Vasconcelos, "Salliency-based Discriminant Tracking", in *proc.of CVPR 2009*, pp.1007-1013, 2009.

[15] H.Grabner, C.Leistner, H.Bischof, "Semi-Supervised On-Line Boosting for Robust Tracking", in *Proc. ECCV 2008*, pp.234-247, 2008.

[16] T.Woodley, B.Stenger, R.Chipolla, "Tracking using Online Feature Selection and a Local Generative Model", in *Proc. BMVC 2007*, 2007.

[17] K.Nishida, T.Kurita, M.Higashikubo, "ONLINE SELECTION OF DISCRIMINATIVE PIXEL-PAIR FEATURE FOR TRACKING", in *Proc. SPPRA 2010*, 2010.